



Privacy in Socially Intelligent ICT

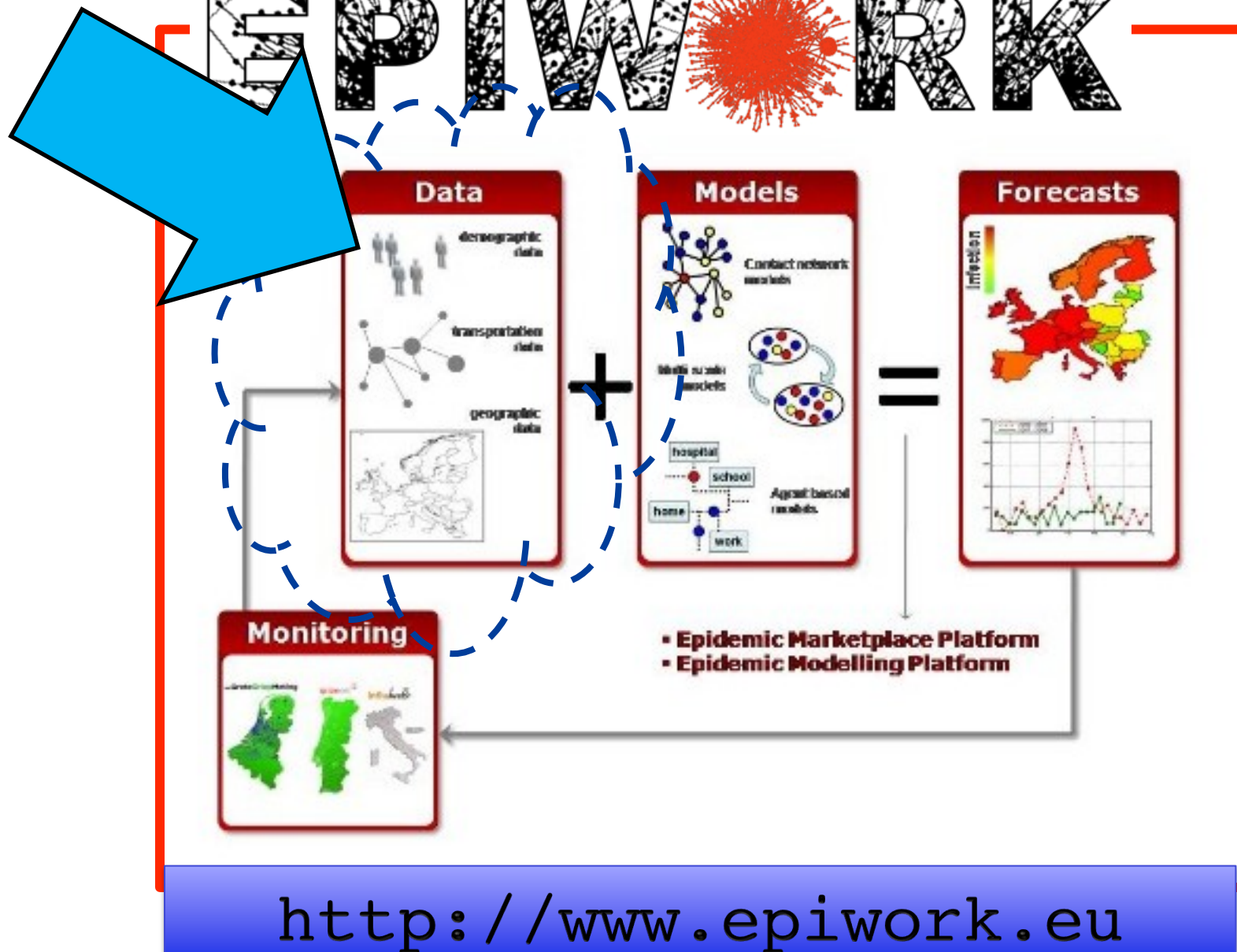
Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências,
Departamento de Informática

mjs@di.fc.ul.pt

Towards a Science of Socially Intelligent ICT
Imperial College London, 3rd August 2010

EPIWORK



[Home](#) [Blog](#) [Data](#) [SPARQL](#) [Apps](#) [Ideas](#) [Forum](#) [Wiki](#) [Resources](#) [About](#)[Home](#) › [Search Data Feeds](#)

Subscribe by **RSS**

Community
Log in / Sign up

Your search returned 5 records

Alcohol is linked to many disease conditions and disease in established market economies. The air
Tags: health-well-being-and-care alcohol nhsscoti
 morbidity health-and-social-care attributable-fracti

Tags: [specialist-health-services](#) [well-being-and-conditions](#) [social-care](#) [nhs](#)

 MONDAY, NOVEMBER 16, 2009

DATA.GOV

[HOME](#) |
 [CATALOGS](#) |
 [STATE/LOCAL](#) |
 [ABOUT](#) |
 [FAQ](#) |
 [CONTACT US](#) |
 [SUGGEST OTHER DATASETS](#)

DISCOVER.
PARTICIPATE.
ENGAGE.

Search the following Data.gov catalogs:



FEATURED TOOL:

This Medicare and Medicaid Statistical tool offers approximately 300 pages of statistical information about Medicare, Medicaid, and other Centers for Medicare & Medicaid Services (CMS) programs. The Supplement includes charts and tables showing health expenditures for the entire U.S. population, characteristics of the covered populations, use of services, and expenditures under these programs. It is one of the most comprehensive sources of information available on health care finance in the U.S.

**CENTERS for MEDICARE &
MEDICAID SERVICES**



[VIEW THIS TOOL ►](#)

Welcome to Data.gov

The purpose of Data.gov is to increase public access to high value, machine readable datasets generated by the Executive Branch of





How to use Data.gov

Data.gov includes searchable **data catalogs** providing access to data in three ways: through the "raw" data catalog, the tool catalog

Número de participantes

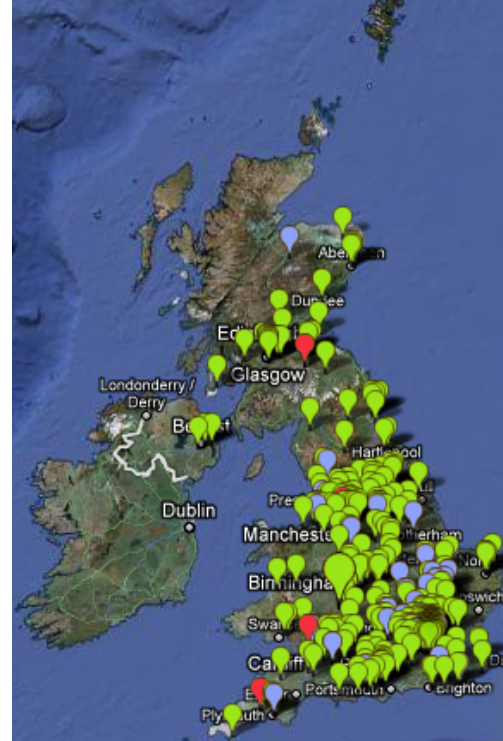
Portugal		5477 participantes
Holanda		20906 participantes
Bélgica		5839 participantes
Itália		3019 participantes
México		3950 participantes
Brasil		370 participantes
Reino Unido		5551 participantes
Australia		7776 participantes
Montreal (CA)		951 participantes

de **Grote Griep Meting**.nl

-  neting (15444)
-  verkouden (3602)
-  vermoedelijk griep (87)
-  grens postcode gebied



2 Aug 2010 - Assyst, London



gripenet 

<http://www.gripenet.pt/>

INFLUWEB

 persone senza sintomi  persone con raffreddore  persone con sintomi



Other Internet Monitoring Sources



Epidemiologists across the world are seizing the opportunity of tapping OSNs

disease	location	source	author	evidence	score	date
H1N1	India	Twitter	SwineFluNewsIN (Ken Wright)	Asia: Mass swine flu vaccination in China as rise time approaches http://tinyurl.com/yjbymbn #swineflu #h1n1	0	2009-10-23 12:54:19
H1N1	India	Twitter	IndyaNews (India News)	H1N1 kills 4 more, toll mounts to 431 http://bit.ly/2GrSH6 #News	0	2009-10-23 11:45:11
H1N1	India	Twitter	SwineFluNewsIN (Ken Wright)	Beware of Online Swine Flu Cures http://tinyurl.com/yk6m322 #swineflu #h1n1	0	2009-10-23 11:42:21
H1N1	India	Twitter	praisid (Praisid Siroraj)	H1N1 kills 4 more, toll mounts to 431 http://bit.ly/XLfkC	0	2009-10-23 11:38:12
H1N1	India	Twitter	SwineFluNewsIN (Ken Wright)	Swine flu cures still online notwithstanding supervision crackdown http://tinyurl.com/y1755c5 #swineflu #h1n1	0	2009-10-23 11:19:30
H1N1	India	Twitter	SwineFluNewsIN (Ken Wright)	Swine Flu Portugal News http://tinyurl.com/pgyyhs #swineflu #h1n1	0	2009-10-23 10:09:26
H1N1	India	Twitter	SwineFluNewsIN (Ken Wright)	2 Aug 2010 - Asyst, London 'advantageous thing' http://tinyurl.com/yzhm97e #swineflu #h1n1	0	2009-10-23 09:42:33

Researchers have professional and legal responsibilities to their respondents that are embodied in the procedures of a research study. Underlying these specific responsibilities are **four fundamental ethical principles**:

2 Aug 2010 - Assyst, London

Respondents should be:

1. willing participants in survey research;
2. appropriately informed about the survey's intentions and how their personal information and survey responses will be used and protected;
3. sufficiently satisfied with their survey experience;
4. willing to participate again in survey research.

Other Internet Monitoring Sources

google.org Flu Trends

[Google.org home](#)

Flu Trends

Japan

[Download data](#)

[Home](#)

[How does this work?](#)

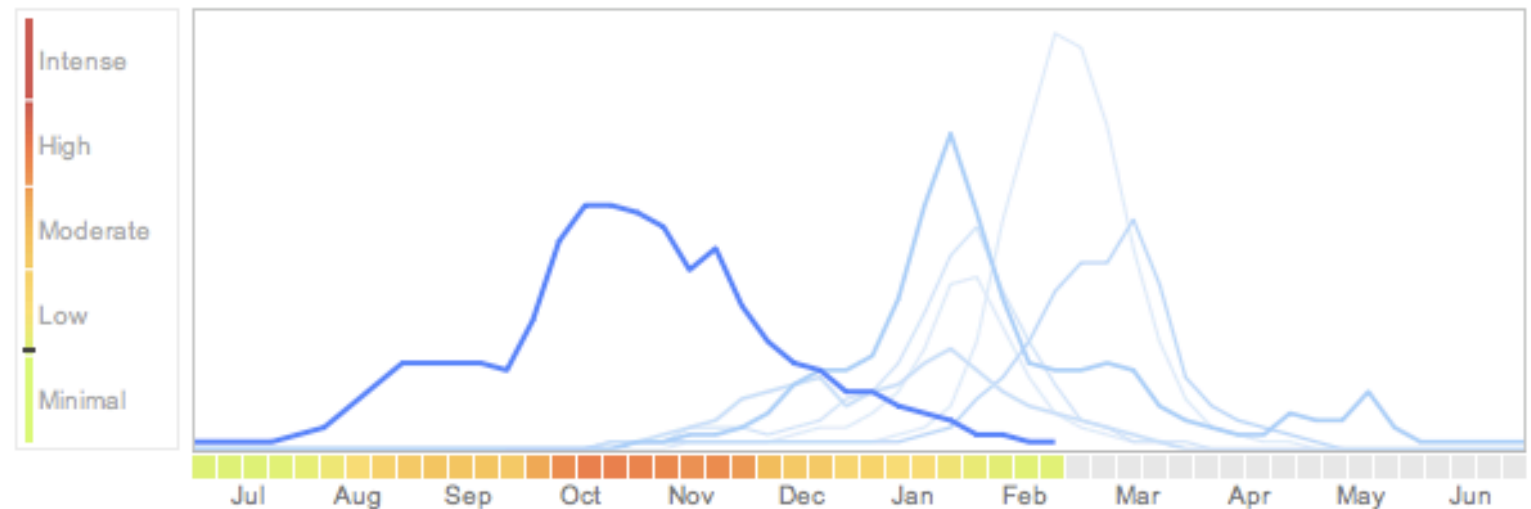
[FAQ](#)

Explore flu trends - Japan

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

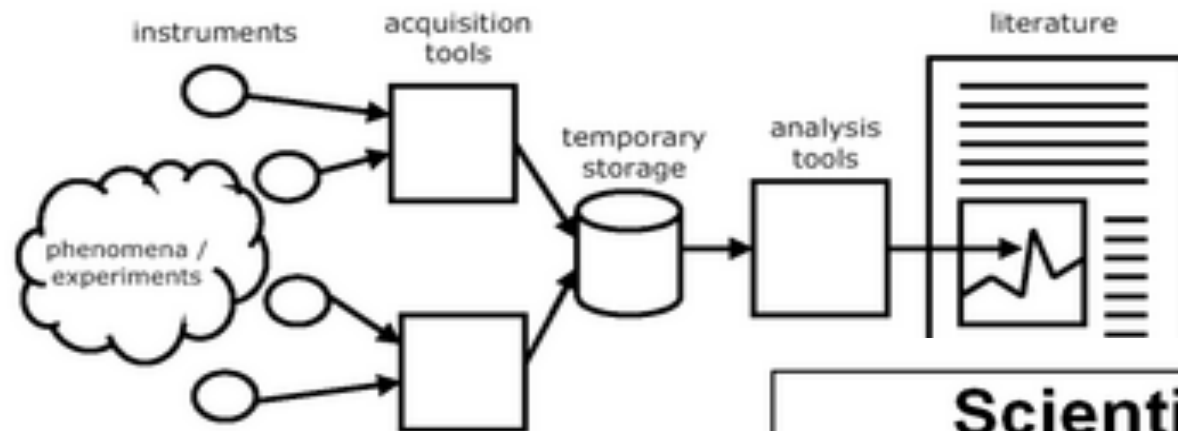
National

● 2009-2010 ● [Past years ▼](#)

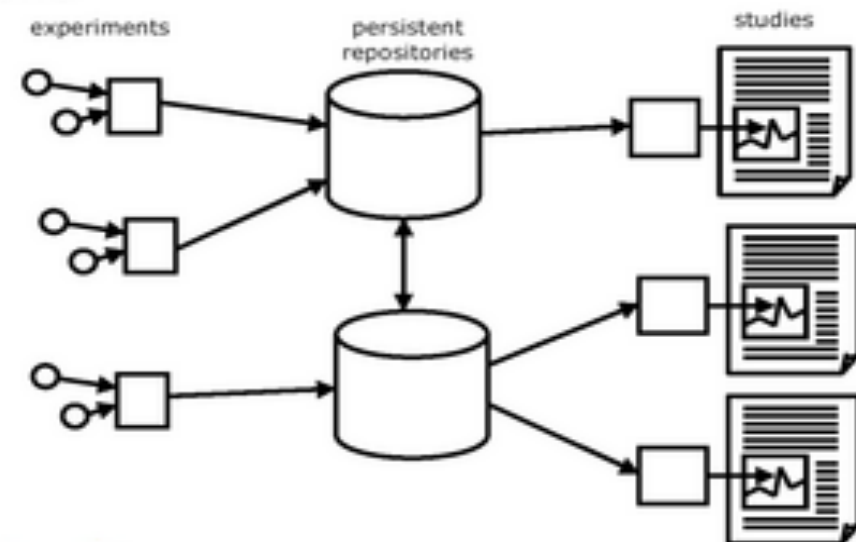


Scientific Data Lifecycle

From this (publish and forget)...



...to this one



OK, let's assume we're about to develop a new SIICT System that will look into OSNs and create a shared repository.

How do we proceed?

- We inform the users we'll start scanning their UGC and their friends before deriving any conclusions about them (their group)
- We just scan the data, because it is public.
- Do we publish/share the scanned data? Under what access terms?

What should we do with the collected data on social behaviours?

- There are two options:
 - A: the Google Flutrends option: see how beautiful are the charts I can show to you
 - B: the e-science approach

Archiving OSN data

- The entire twitter database is being copied to the LOC!



LIBRARY OF
CONGRESS

Anonymization in OSN ?

- While companies tend to focus on the removal of personally-identifiable information (PII), the studies show that re-identification can occur even by combining non-PII, such as movie ratings in the Netflix study or search engine queries in the AOL example.



- **Netflix:** Arvind Narayanan and Vitaly Shmatikov
 - Robust De-anonymization of Large Sparse Datasets
 - De-anonymizing Social Networks

The Incompatibility Issue

- Privacy (as we understand it today) and social nets (as we use them) are incompatible.
- **Danah boyd:** “Putting Privacy Settings in the Context of Use”
 - See "Project Gaydar"

Privacy technologies in IT

- Policy Communication
 - **P3P** - The Platform for Privacy Preferences. P3P is a standard for communicating privacy practices and comparing them to the preferences of individuals.
- Policy Enforcement
 - **XACML** - The Extensible Access Control Markup Language together with its Privacy Profile is a standard for expressing privacy policies in a machine-readable language which a software system can use to enforce the policy in enterprise IT systems.
 - **WS-Privacy** - "Web Service Privacy" will be a specification for communicating privacy policy in web services. For example, it may specify how privacy policy information can be embedded in the SOAP envelope of a web service message.

This is no solution!

Privacy Laws & Regulations

- EU data protection directive, etc

Any SIICT system will have a database or access one with “personally identifiable” information

- Will we ask permission from the national data protection committee and each and every user for each new query?

On the nature of SIICT Systems

- Problem is not the network, but the behaviours learnt about individuals from the network
- No Silver Bullet
- “Classic” Engineering approaches too slow for 21st century pace
- We are now all part of a huge Living Lab
 - How much longer will the fact that your cat sneezed be relevant?...we might have to ask again.
 - Are we still under control?

Final Reflections on SIICT

- Will the idea of SIICT systems w/ (public?) shared data ever fly?
 - Must be conceived with **privacy in mind from the ground up.**
 - Create socially-inspired design patterns
 - robots.txt for OSN? Privacy by design?
- Setup “Living Labs” for SIICT systems
- Privacy:
 - The mission of scientists is to break the rules
 - Hey, we’re the good guys!